Automatic morphological description of galaxies and classification by an expert system

Monique Thonnat

INRIA

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 387

AUTOMATIC MORPHOLOGICAL DESCRIPTION OF GALAXIES AND CLASSIFICATION BY AN EXPERT SYSTEM

Monique THONNAT

Mars 1985

AUTOMATIC MORPHOLOGICAL DESCRIPTION OF GALAXIES AND CLASSIFICATION BY AN EXPERT SYSTEM

Monique Thonnat

INRIA Sophia Antipolis 06560 Valbonne, France

RESUME:

Nous présentons une méthode de classification automatique de galaxies selon leur type morphologique à partir d'une description qualitative de chaque classe. Le traitement automatique d'une image est effectué en deux phases: des paramètres caractéristiques sont extraits de l'image pour décrire la galaxie, puis ces paramètres sont fournis à un système expert qui détermine le type morphologique de cette galaxie. Les différents algorithmes utilisés au cours de la première phase sont décrits, puis le type d'un système expert spécialisé dans la classification d'objet est montré. Un échantillon de 39 galaxies a été traité; les classes fournies par le système expert sont comparées avec celles obtenues par inspection visuelle des plaques photographiques. La fiabilité des résultats montre que cette technique est bien adaptée à la classification d'objets complexes; la précision de la classification réside dans la quantité d'information contenue dans la base de connaissance.

ABSTRACT:

We propose a computer vision method to classify automatically galaxies into their morphological type according to a qualitative description of each class. The automatic processing of an image of galaxy is made of two stages: first, characteristical parameters are extracted from the image to describe the galaxy; secondly, these parameters are provided to an expert system to determine the morphological type of the galaxy.

The various algorithms used in the first stage are described; then, the design of an expert system specialized in object classification is shown. A sample set of 40 galaxies has been processed; the classes given by the expert system are compared with those obtained by the visual inspection of the plates. The reliability of the results shows that this technique is well-adaptated to the classification of complex objects; the precision of the classification lies in the quantity of information kept in the knowledge base.

TABLE OF CONTENTS

AUTOMATIC MORPHOLOGICAL DESCRIPTION OF GALAXIES AND CLASSIFICATION BY AN EXPERT SYSTEM

M. Thonnat

INRIA Sophia Antipolis
06560 Valbonne, France

1. INTRODUCTION

The classification of galaxies into various morphological types is a very difficult problem, that only few specialists overlook [San61a, Vau59a]. The major difficulty stands in the uniqueness of the direction of sight under which the 3-D object is observed.

Right now, this classification is achieved by the only visual inspection of photographical plates. But, the increasing quantity of data provided by the new instruments (large field Schmidt telescopes, spatial telescopes, high speed digitizers) and the need of objective and explicit reasoning, imply the automation of the classification [Lau84a, Tho84a].

We present an automatic method to classify galaxies, that can be integrated in a data processing program of photographical plates, taken with a large field telescope.

The complete processing of a photographical plate is made of three stages:

1. The plate measuring the brightness of an astronomical field is digitized and the galaxies in it are detected ;

2. For each galaxy detected, specific parameters are extracted;

3. Finally, these parameters are used to classify the processed galaxy; the classification is made with an expert system rather than merely a standard pattern recognition algorithm [Bal82a] ;

The first stage corresponds to two processes (digitization and discrimination of galaxies from stars, other astronomical objects or artifacts) which are well-known in astronomical image processing [Bij81a, Jar81a].

The second and third stages are developed in the following chapters.

2. DESCRIPTION OF THE OBJECTS

2.1. The classification systems

The first classification system has been defined by Hubble in 1926. Since then, some modifications [Hub36a, Vau59a, San61a] have been brought to obtain the most used system: the Hubble revised system. The classification of galaxies

may be described with a continuous label coding T or a hierarchical label coding H.

The morphological type T is an integer value varying between -5 and 10, which characterizes the degree of inhomogeneity of the morphology (figure 1). The lowest value (T = -5) corresponds to a perfect ellipsoïd with an excentricity below 0.7. When T increases, the shape of the galaxy becomes flatter and may be described as the superposition of a small sphere and a surrounding flat disk ($T \leq -1$). The positive values of T correspond to the presence of a spiral structure inside the disk. The extreme value (T = 10) represents a very irregular shape.

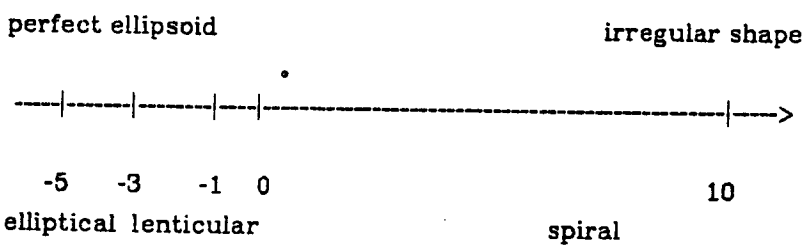

figure 1
The continuous labeling T

The hierarchical label coding H explicits the different patterns present in the galaxy (figure 2a). This code is an alphanumerical chain, where each character represents a special pattern.

- the main character defines the general class of the galaxy: E stands for elliptical, L for lenticular and S for spiral;
- the second character defines the eventual presence of a transversal bar: A stands for a missing bar, B for a present bar and X for the intermediate case;
- the third character precises the shape of the structure in the disk: R stands for a ring, S for an open spiral (or S-shape) and T for the intermediate case;
- other characters precise respectively the continuous type T or some peculiarities in the morphology.

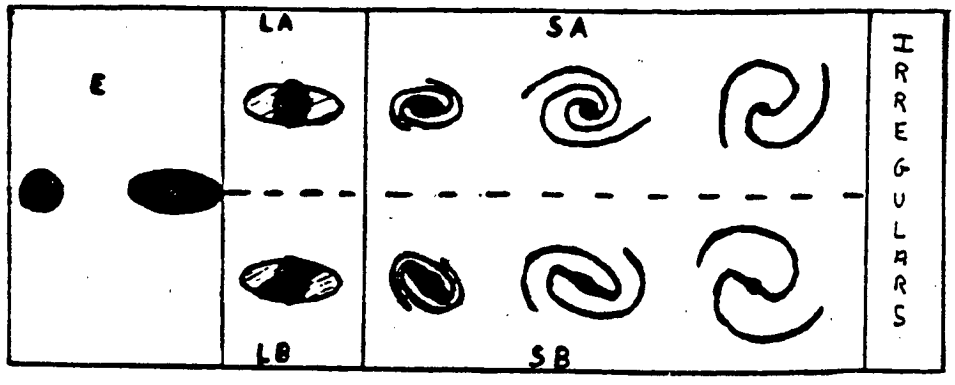

figure 2a
The different structures described by the hierarchical labeling H

For instance (figure 2b):

LA represents a lenticular galaxy (L), with no bar (A);

SBR3 represents a spiral galaxy (S) with a bar (B), an internal ring (R), and a type T equal to (3).

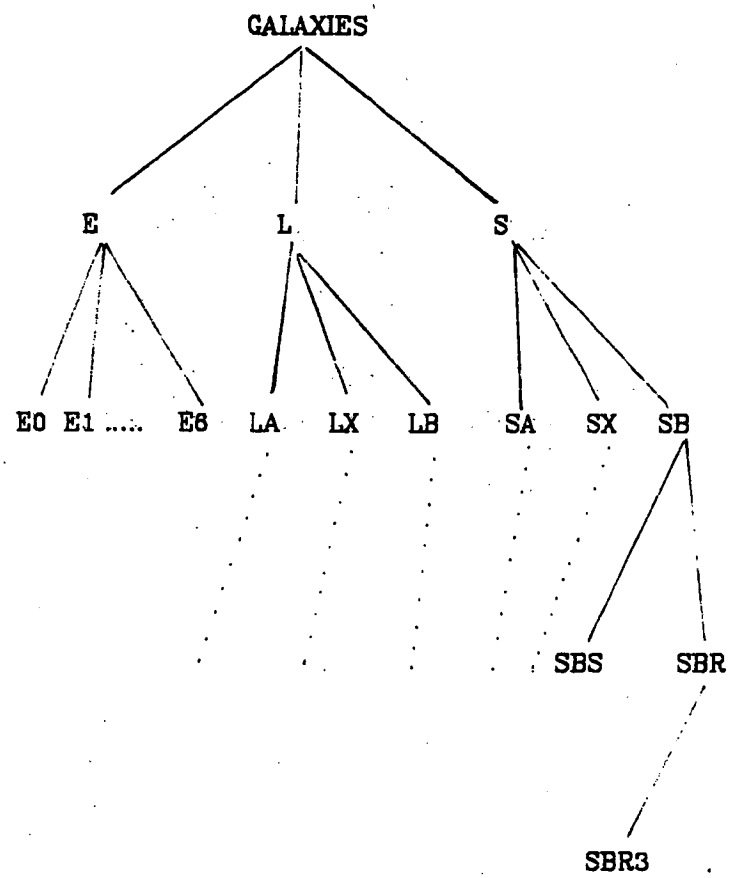

figure 2b
The hierarchical label coding H

2.2. Examples of images

Figure 3 shows some examples of typical images of galaxies provided by the digitization and detection stage for the processing of a photographical plate. These galaxies of the Virgo cluster have been observed by J-D Strich with the Schmidt telescope at *CERGA**. The sampling rate is (20 * 20) microns, and the image size is (512 * 512) pixels.

CERGA: Centre d'Etudes et de Recherches Geodynamiques et Astronomiques

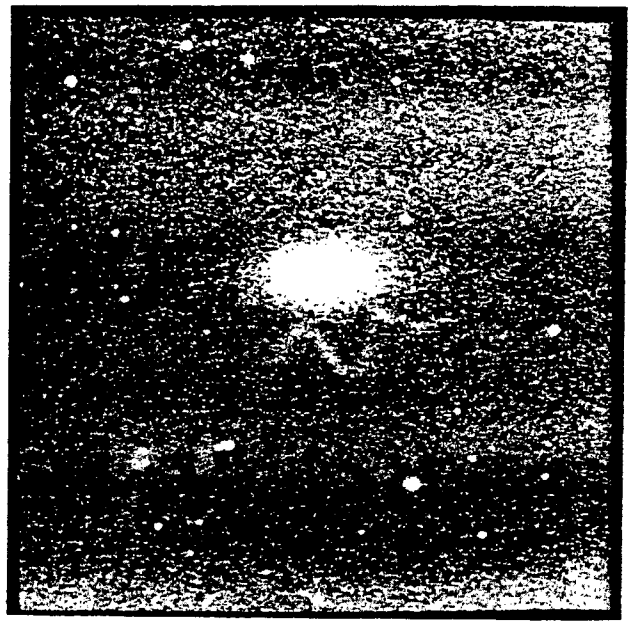

figure 3a
An elliptical galaxy (NGC 4473)

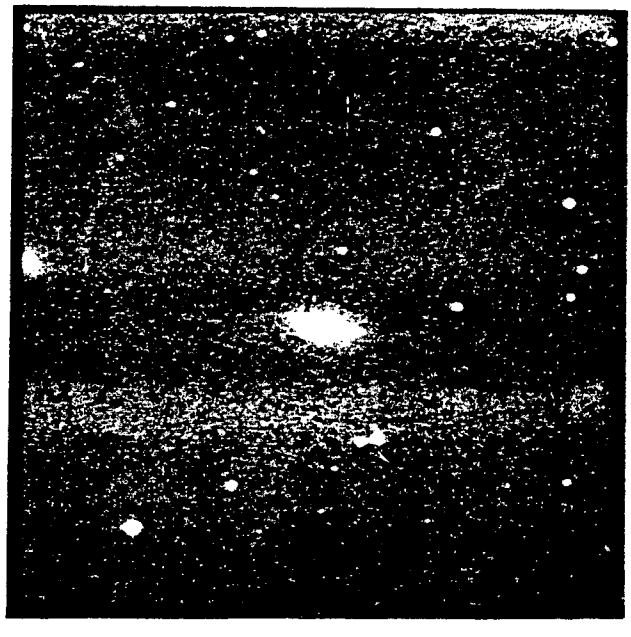

figure 3b
A lenticular galaxy (NGC 4474)

figure 3c
A spiral galaxy (NGC 4569)

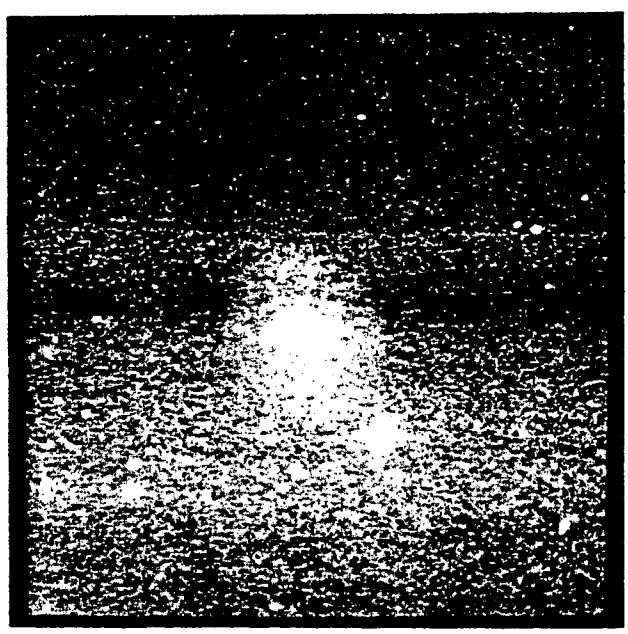

figure 3d
A late spiral galaxy (NGC 4571)

Among the properties of such images, we have to notice that:

* These images have a very low signal to noise ratio, which becomes critical for faint objects.
* There is no clear limit discriminating the galaxy from the background (see figure 4).

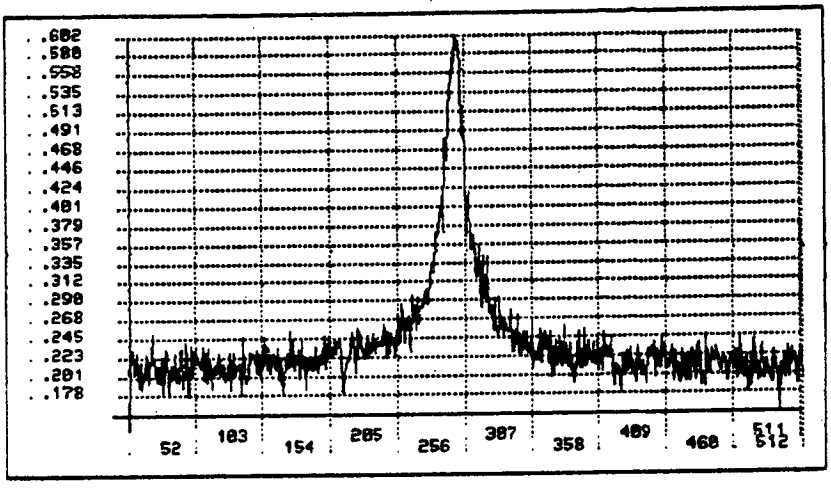

figure 4
plot of a line passing accross the center of NGC4473

* Other objects may be close, or even partially overlapping the galaxy.

3. PREPROCESSING

3.1. The different stages

All these mentioned properties show that a preprocessing phase is necessary to enhance the quality of the image before parameters can be easily extracted from the isolated galaxy. This phase must perform the following functions:

- locate the exact position of the galaxy;
- define the limits of the galaxy;
- build a cartography of the background;
- eventually substract the objects in the neighbourhood;

- eliminate the noise;

The data processing techniques needed for these functions are strongly dependent on the values of the images; although, as the system of classification must be completly automatic, even in this preprocessing phase, we must avoid all interactive methods, (in particular manual selection of thresholds) and choose the most adaptive techniques.

3.2. The algorithms

3.2.1. Location of the position

Whatever may be the value of the morphological type T, the location of the galaxy is defined by the position of the center of the spheroïdal bulge. The bulge is the brightest region in the galaxy and can be modelized by a 2-D gaussian distribution. So, the position is obtained by locating the maximum of the correlation of the image and a model of the bulge.

This "center" differs from the geometrical center in the case of an asymmetry in the spiral morphology of S galaxies, making irrelevant the measure of the centre of gravity.

3.2.2. Limits of the galaxy

A galaxy is a nebulous object, with a global intensity decreasing continuously from the bulge. The precise position of the bulge being known, a curve showing the variation of the average intensity from the bulge in terms of the radius, is built (figure 5). We take as extremal radius of the galaxy the radius which corresponds to the asymptotic value of this curve (f). Then, we benefit of this knowledge of the limits of the galaxy to reduce the amount of data and thus optimize the computing time required for processing.

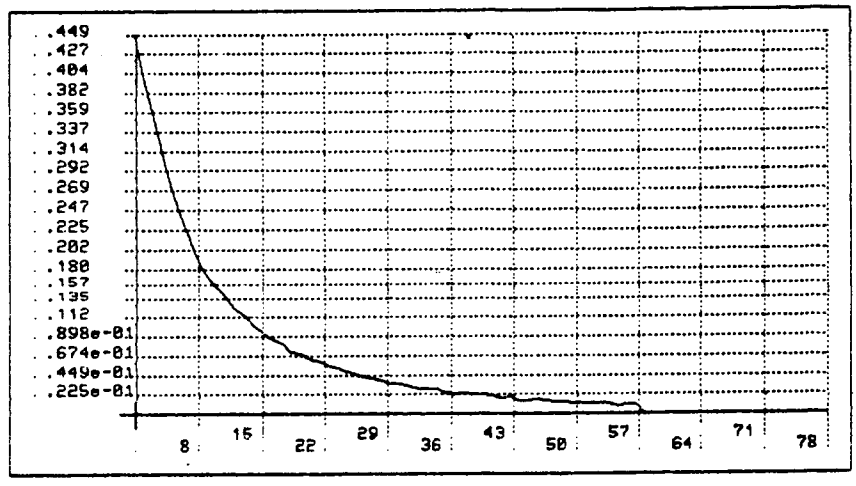

figure 5
Radial variation of the intensity from the bulge (NGC4473)

3.2.3. Background substraction

The measured intensity in the region of the galaxy is the superposition of the brightness of the galaxy and the intensity of a background. The background is negatively defined by opposition to the object of interest. In the circumstances, the background is made of the chemical fog, the unresolved objects and possibly the galaxies cluster background.

The brightness of this background varies in the image, so a complete cartography of the background is needed in order to substract it from the initial image. The values of the background though being non uniform have smooth continuous variations. Thus, a method based on the interpolation of the intensities in the background regions is available, and has been implemented.

First, the image is divided into elementary cells; the size of these cells grows with the area of the image (for instance, the cells have an area of 16*16 pixels for an image of 512*512 pixels). The cells which have no pixels inside the limits of the galaxy, or with an intensity corresponding to a star (see the next section), are labeled as background cells. The cartography of the background inside the limit of the galaxy, is built using a linear interpolation.

3.2.4. Star removal

As the brightness of the galaxy decreases strongly from the central bulge, all the cells outside the bulge with an intensity greater than half the central intensity (Im) are labeled as stars. The limit of the bulge is determined from the curve of the radial intensities by the radius R_{limit} defined by :

$$I(R_{limit}) - f = \frac{1}{3} \times (Im - f)$$

3.2.5. Noise filtering

Once the background and the stars have been removed from the image restricted to the galaxy, the next preprocessing stage is noise elimination. We choose the median algorithm with a 3*3 window size for its property of preserving the edges.

Figure 6, 7, 8 and 9 show the results of the preprocessing on the galaxies previously described.

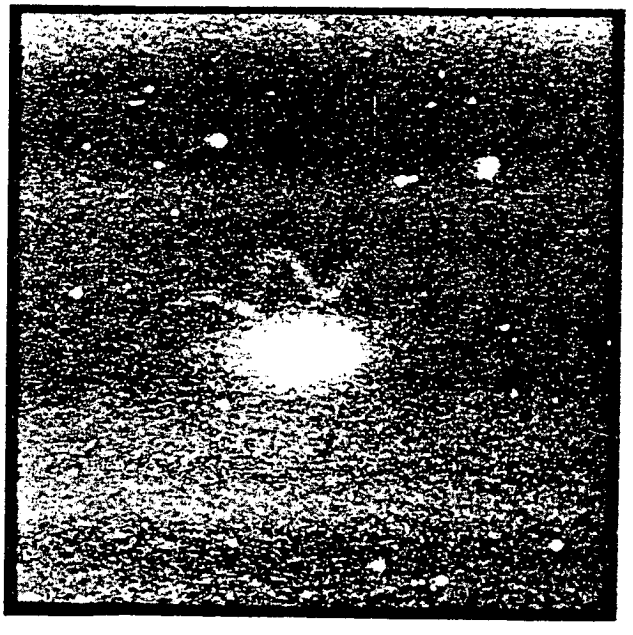

figure 6a
Galaxy NGC4473

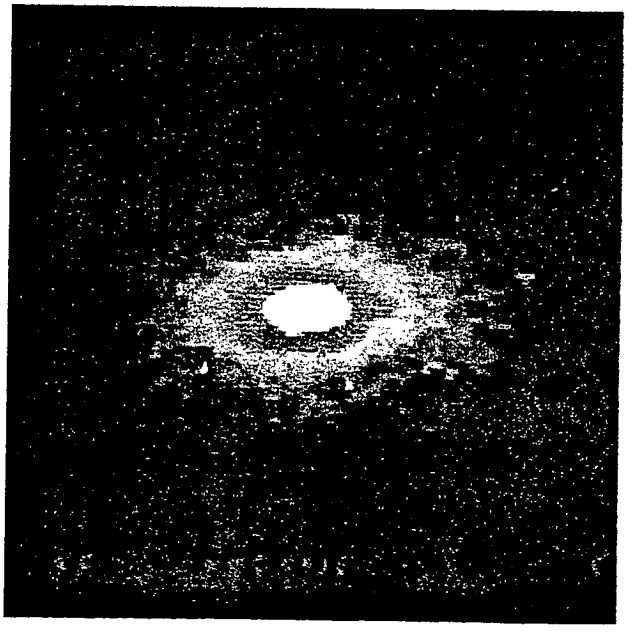

figure 6b
The same galaxy after preprocessing

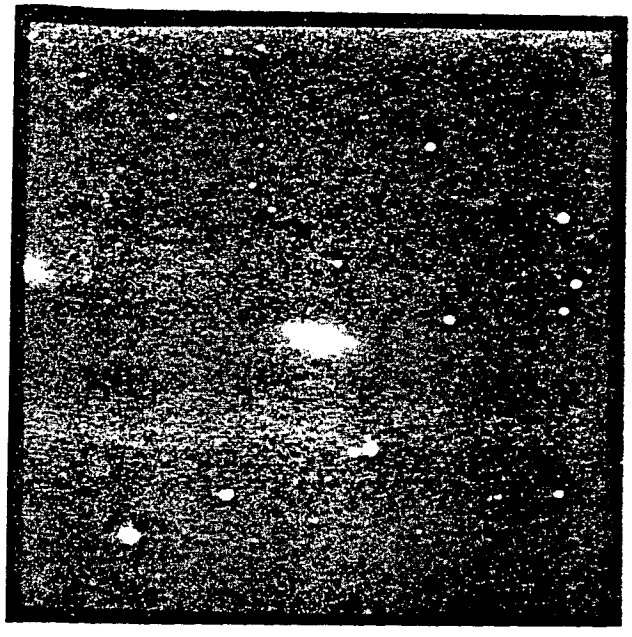

figure 7a
Galaxy NGC4474

figure 7b
The same galaxy after preprocessing

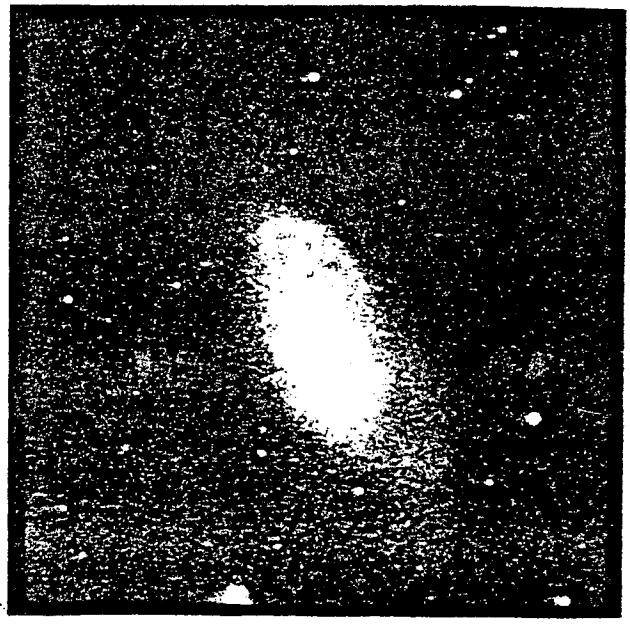

figure 8a
Galaxy NGC4569

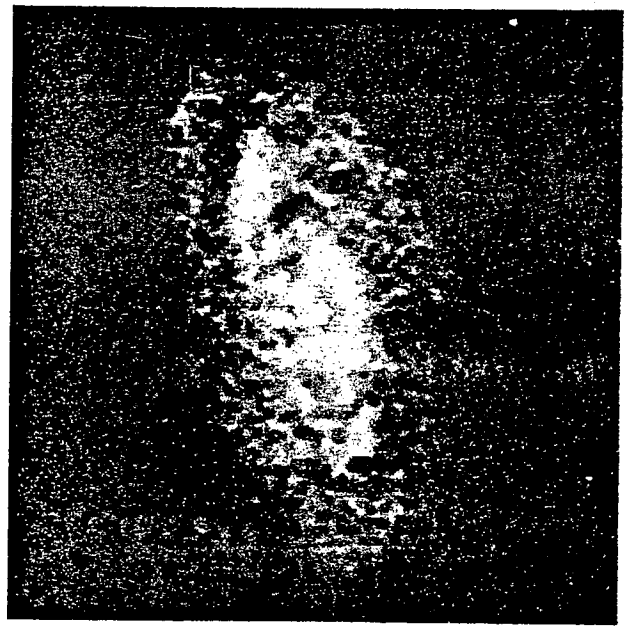

figure 8b
The same galaxy after preprocessing

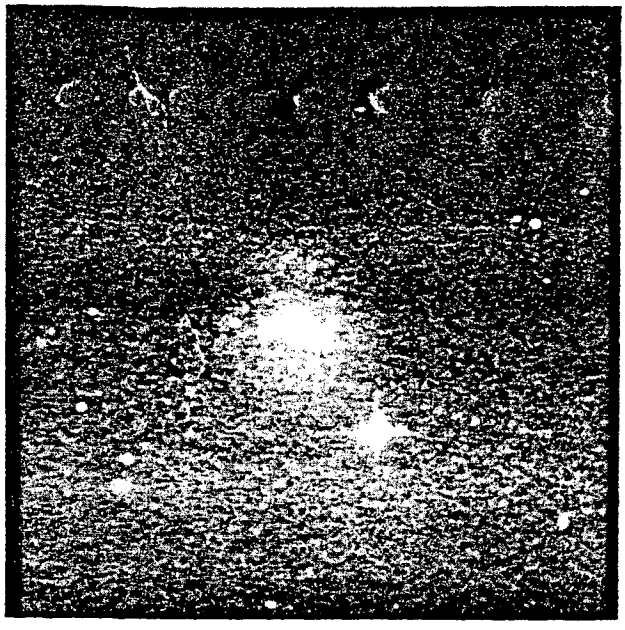

figure 9a
Galaxy NGC4571

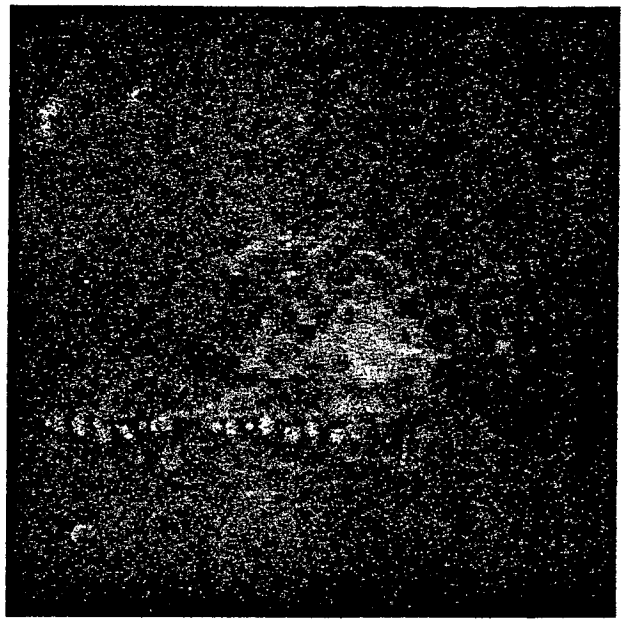

figure 9b
The same galaxy after preprocessing

4. PARAMETER EXTRACTION

4.1. Introduction

In this section we present the various parameters extracted from the galaxies to describe their shape.

4.2. Principal axis

The first parameter computed is the orientation of the principal axis of the galaxy. The value of its direction is given by the angle Θ which minimizes the moment of inertia J :

$$J = \sum_{x,y} d(x,y) \times (y\cos\Theta - x\sin\Theta)^2$$

Let Θ_1 and Θ_2 be the two solutions of

$$\tan(2\Theta) = 2\frac{\sum_x \sum_y d(x,y)xy}{\sum_x \sum_y d(x,y)(x^2-y^2)}$$

with $J(\Theta_1) < J(\Theta_2)$

Θ_1 is the orientation of the principal axis;

Θ_2 is the orientation of the small axis (axis orthogonal to the principal axis).

A set of curves representing the intensity of the galaxy from the center in different directions are built. These directions are the two principal half-axis and the two small half-axis. Figure 10 displays the shape of these curves for the elliptical galaxy NGC4473 and the spiral galaxy NGC4569.

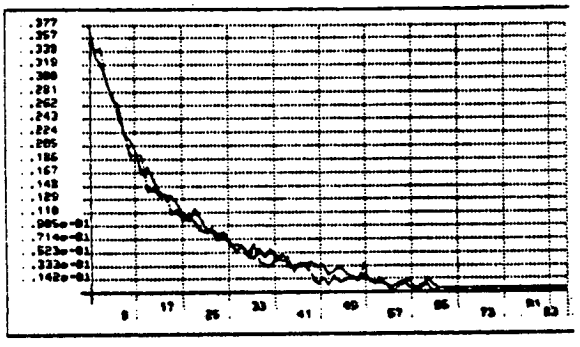

figure 10a
Distribution of the luminosity from the center of NGC4473 in the Θ_1 and $\Theta_1+\pi$ directions

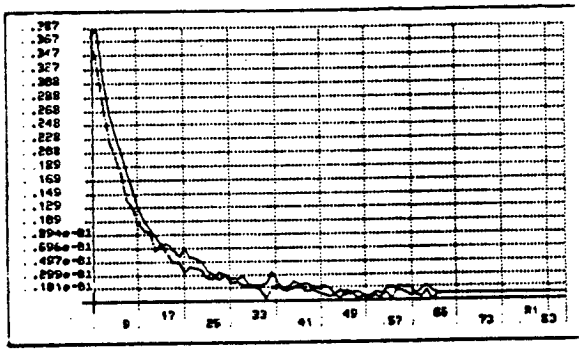

figure 10b
Distribution of the luminosity from the center of NGC4473 in the Θ_2 and $\Theta_2+\pi$ directions

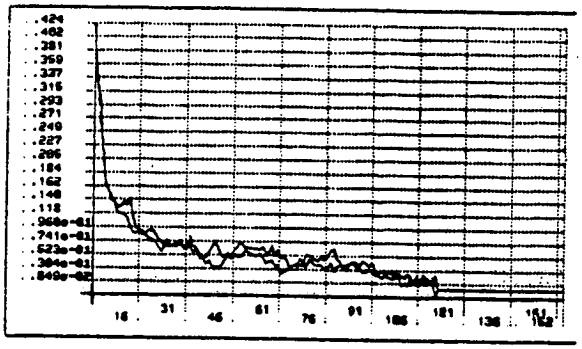

figure 10c
Distribution of the luminosity from the center of NGC4569 in the θ_1 and $\theta_1+\pi$ directions

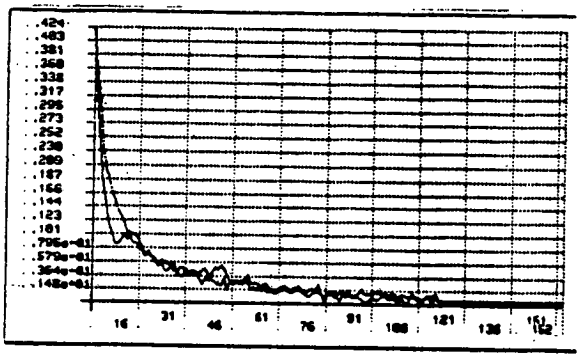

figure 10d
Distribution of the luminosity from the center of NGC4569 in the θ_2 and $\theta_2+\pi$ directions

4.3. Ellipticity

The shape of the galaxy is very dependent on the view-point under which the object is observed. Therefore, an important measure is the apparent ellipticity of the galaxy:

$$e = 1-\frac{b}{a}$$

with a and b respectively the width of the principal axis and the width of the orthogonal axis.

This value is provided by the previous curves indicating the width of the galaxy in the principal axis and in the small axis directions. As these curves are noisy and have a smooth slope due to the diffusion in the emulsion, we use an integrated width value. Let F(x) be the average distribution of the intensity along the principal axis, and f(x) the average distribution of the intensity along the small axis; let w_1 and w_2 respectively the width of the galaxy in the two directions:

The widths are deduced from:

$$\int_0^{w_1} F(x)dx = \frac{3}{4}\int_0^{\infty} F(x)dx$$

$$\int_0^{w_2} f(x)dx = \frac{3}{4}\int_0^{\infty} f(x)dx$$

Finally the apparent ellipticity is given by:

$$e = 1 - \frac{w_2}{w_1}$$

4.4. Size

In order to estimate the reliability of the measured parameters we compute the size of the galaxy S with the hypothesis of an elliptical shape:

$$S = \pi w_1 w_2$$

4.5. Average profile

Using the measure of the apparent ellipticity e we build the curve of the average distribution of the intensity from the center g(x).

The curve g(x) is the radial distribution of the intensity in the new referential (X, $Y = \frac{w_1}{w_2} \times y$), with X the coordonate on the principal axis and Y the coordinate on the small axis .

4.6. Projected profile

Theorical studies [Vau76a, Wat82a] have shown that the radial variation of the luminosity from the center of the galaxies is the sum of two functions;

The first function is the spheroïdal component characterizing the elliptical galaxies.

- The intensity I is given in terms of the radius r by:

$$I_1(r) = a_1 e^{-b_1 r^{\frac{1}{4}}}$$

- The density $d = \log I / I_0$ is given in terms of the radius by:

$$d_1(r) = \alpha_1 - \beta_1 r^{\frac{1}{4}}.$$

The second function is the flat component characterizing the disk of the lenticular and spiral galaxies.

- The intensity I is given in terms of the radius r by:

$$I_2(r) = a_2 e^{-b_2 r}$$

- The density $d=\log I/I_0$ is given in terms of the radius by:

$$d_2(r)=\alpha_2-\beta_2 r$$

In order to avoid the influence of the viewpoint under which the galaxy is observed we build a new curve: the projected profile d_p. This curve is obtained by orthogonal projection of the densities d of the galaxy along the principal axis.

$$d_p(u) = \sum_{\substack{x,y \\ u=x\cos\theta+y\sin\theta}} d(x,y)$$

With θ the direction of the principal axis, x, y the coordinates of the pixels belonging to the galaxy.

Figure 11 shows the shape of this projected profile for an elliptical galaxy NGC4473 and a spiral one NGC4571.

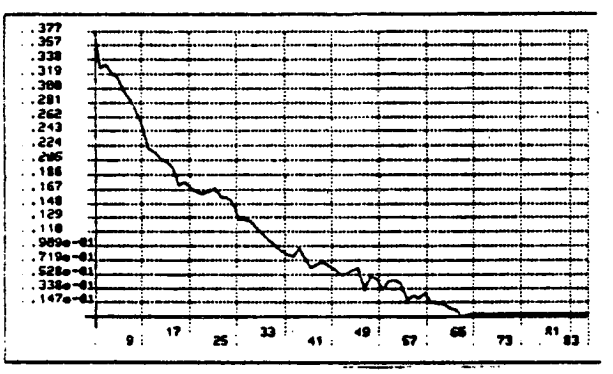

figure 11a
Projected profile of NGC4473

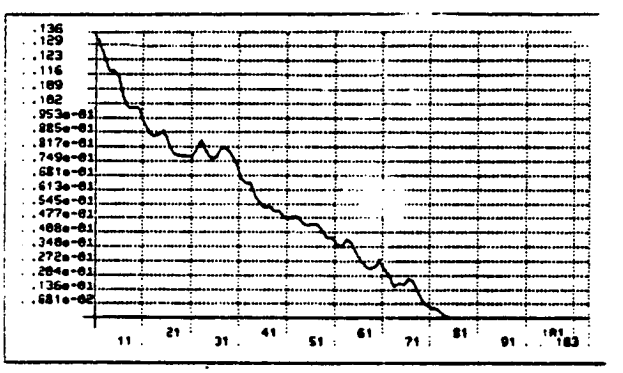

figure 11b
Projected profile of NGC4571

From this curve we extract two parameters:

- the parameter (profile) which is the ratio of the mean square errors made by approximating the projected profile respectively with d_1 and d_2 for r greater than the radius of the bulge. The estimation of the radius of the bulge is obtained from the average profile previously built.

$$profile = \frac{\sum_{r=r_{bulge}}^{r_{limit}} (d_p(r)-d_2(r))^2}{\sum_{r=r_{bulge}}^{r_{limit}} (d_p(r)-d_1(r))^2}$$

- the parameter (linear-err) measuring directly the mean square error made by approximating the complete projected profile for r=0 to $r=r_{max}$ with a linear function.

$$linear-err = \sum_{r=0}^{r_{limit}} (d_p(r)-d_1(r))^2$$

4.7. Contours

4.7.1. Contour building

In order to describe the variation of the structure in the different regions of the galaxy, we need to extract several isophotes. The isophotes must be completely representative of each region in the galaxy and they must be built without interactive process. First we compute five thresholds from the distribution of the density along the principal axis F(x), then these thresholds are used to obtain five binary images on which we apply an edge detector algorithm.

The five thresholds $t_i, i \varepsilon [1, \ldots, 5]$ are given by:

$$\int_0^{t_i} F(x)dx = \frac{1}{n_i} \int_0^{\infty} F(x)dx$$

with n_i respectively equal to 0.20, 0.50, 0.75, 0.85 and 0.90

For each thresholded image, we use a Sobel edge detector, then we exhibit the maximal chain issue from the contour chaining algorithm of the INRIMAGE [Cip84a] library.

4.7.2. Examples of contours

Figure 12 shows the contours associated with the four previously mentioned galaxies:

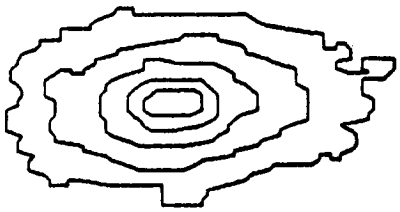

figure 12a
Contours extracted from NGC4473

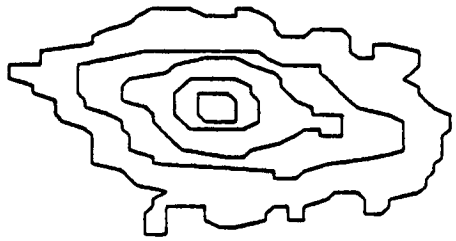

figure 12b
Contours extracted from NGC4474

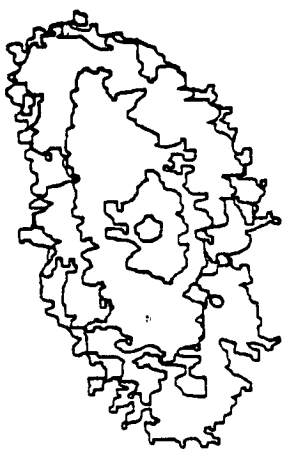

figure 12c
Contours extracted from NGC4569

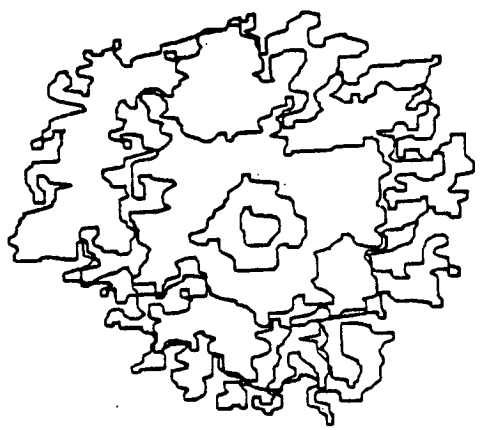

figure 12d
Contours extracted from NGC4571

4.7.3. Parameters extracted from these contours

From each of these contours we extract five parameters characterizing its shape: ellipticity, angle of the principal axis, relative position of the center, compacity and distance of the closest ellipse.

the angle of the principal axis

In the same way as for the image, we compute the angle Θ_i which minimizes the inertial moment J_i and the orthogonal angle ψ_i.

the ellipticity

For each contour we compute the ellipticity (the description has already been made).

the relative position of the center

For each contour we compute the euclidian distance (center–err) between the center of the bulge and the center of gravity of the contour g_i.

the compacity

The compacity C measures the roundness of a shape and is minimal for a circle ($c \leq 4\pi$).

$$c = \frac{Perimeter^2}{area}$$

In fact we take into account the value of the estimated ellipticity $e = 1 - \frac{b}{a}$ and we measure the quantity:

$$C = \frac{b}{a} \times \frac{c}{e\pi}$$

the distance of the closest ellipse

The last parameter (ellipse–err) is the distance between each contour and the ellipse which as its main axis in the Θ_i direction, an ellipticity e and is centered in the center of gravity g_i.

Normalization is performed as follows:

Let S_E the area of the ellipse, S_C the area of the contour and S_{dif} the sum of the areas between the two closed curves

$$ellipse\text{–}err = \frac{S_{dif}}{S_E + S_C}$$

Sd is approximated by dividing the area into small triangles, the vertices of which are respectively, the points of the contour C_i, the projections of these points on the ellipse p_i and the intersections of the contour and the ellipse P_j (figure 13).

Let x_{p_i} and y_{p_i} the coordinates of the projections p_i and x_i and y_i the coordinates of the points C_i belonging to the contour.

$$x_{p_i} = a.\cos\left(arctg\frac{ay_i}{bx_i}\right)$$

and

$$y_{p_i} = b.sin\left(arctg\frac{ay_i}{bx_i}\right)$$

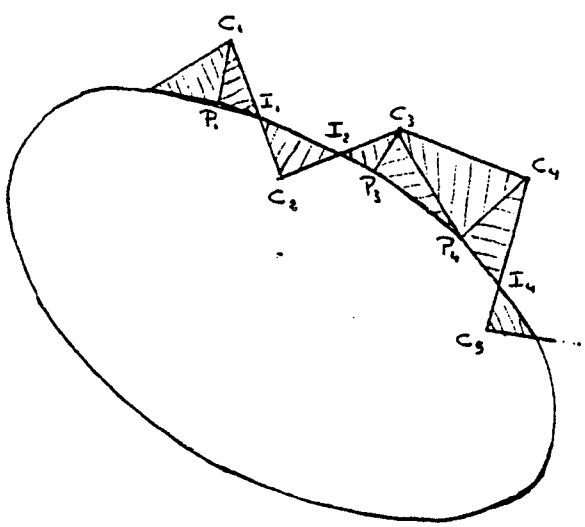

figure 13
Approximation of the area S_{dif}

4.1. Example of measured parameters

We display the parameters extracted from NGC4569.

orientation : 79.33
ellipticity: 0.50
linear-err : 0.039
profile : 2.77
area : 38013.3

contours:	center-err	ellipse-err	compacity	angle	ellipticity
c1 :	1	0.06	1.6	-6.3	0.09
c2 :	2	0.21	5.4	70.8	0.36
c3 :	11	0.16	7.1	83.6	0.58
c4 :	9	0.21	11.8	76.8	0.47
c5 :	7	0.14	8.6	76.1	0.50

5. AUTOMATIC CLASSIFICATION

5.1. Choice of the method

In section 2. we have seen that the various 3-D models of classes are not isotropic. Now, the images of galaxies to classify are the 2-D projections of the real objects on an astronomical plate. So, for each 3-D model we observe a great diversity in the shape of the digitized images.

The main problem of this classification is to match a 2-D image with a 3-D model. This problem cannot be expressed as a graph isomorphism problem and standard pattern matching algorithms [Pav77a] are not available in this case. Actually, we have to valid hypothesis with criteria similar to these ones used by the experts. (for instance, if the apparent excentricity of a galaxy is great, this galaxy itself is flat and seen on the side, so the type T is greater than -5). So, as we need a tool which both works on the descriptions of the classes and on these criteria, we have developed an expert system.

Artifial intelligence methods essentially provide easy symbolic manipulation by means of languages as Lisp or Prolog (or object-oriented languages), as well as the possibility of using heuristics to decrease the complexity of the problem.

Expert systems stand among the most attractive tools recently develcped in the A.I. community [Nil82a, Hay83a]. They have often been applied to process a huge amount of data (for example Mycin) [Sho76a] for which, in addition to the already mentioned properties, they offer nice properties of modularity and extensibility. Actually, they present a complete separation of the knowledge

from the control structure and the addition of new rules in the knowledge base does not need other modification in the system.

In the computer vision domain, several systems [Slo79a, Bro81a, Han78a] using this approach have been proposed.

The expert system we present in the following section has been specially studied to solve the problem of the object classification.

5.2. The knowledge base

Previous studies [Gra84a, Tho85a] have shown that a classical rule interpreter is not adaptated to classify objects into well-known classes. Actually, to focus the reasoning on the classification of an object into a class, we need to introduce in the knowledge base the explicit descriptions of the various models to enable the estimation of a distance between the object and the possible models. Moreover, the knowledge base must contain rules; precisely, the production rules formalism expresses in an explicit way the subjective criteria used by the experts, and allows the passage of the quantified parameters to symbolic descriptors.

So, a system using both productions rules and frame-like objects (prototypes) to describe models has been developped. It has been implemented in an object oriented language, CEYX, [Hul83a] an extension of LeLisp [Cha84a] which is an efficient language to build data structures like production rules or frames. An extensive description of this system can be found in [Gra85a].

5.2.1. The symbolic parameters

The description of a class is made by the astronomer in terms of symbolic parameters which represent the different structural patterns of the galaxies. We have chosen to introduce such parameters in the knowledge base in addition to the quantified parameters previously described.

The symbolic parameter H

This symbolic parameter corresponds to the hierarchical label coding; it takes the values of the possible classes:

H: E, L, S, E0, E1, ..., E6, LA, LB, LA-, LB- ,LA+ ,LB+ , SA, SB, SA0, ...SA7, SB0, ..SB7

As the irregular galaxies are not yiet classified the associated values are not available.

The symbolic parameter T

This parameter corresponds to the continuous label coding; though it takes numerical values, it is not a measured parameter, but each value represents a label:

T: -5, -3, -2, -1, 0, 1, ...,7

The values greater than 7 are not considered as they correspond to irregular galaxies.

The symbolic parameter bar

This parameter specifies the knowledge about the possible presence of a transversal bar; so, the symbolic values are:

bar: present, absent, unknown

The symbolic parameter shape

This parameter qualifies the general structure of the galaxy; it may take the values:

shape: elliptical, average, spiral

The symbolic parameter isophotes

This parameter points out the degree of pertubation of the contours; the possible values are:

isophotes: smooth, normal, distorted

The symbolic parameter arms

This parameter qualifies the arms of the galaxie;the values are:

arms: absent, incipient, evident, branched

The symbolic parameter bulge

This parameter mentions if the central bulge can be detected on the projected profile; so, the values are:

bulge: visible, unvisible

The symbolic parameter flatness

This parameter qualifies the apparent ellipticity of the galaxy.

flatness: null, negligible, very-faint, faint, light, average

The symbolic parameter centring

The parameter centring measures the difference between the centers of the galaxy and the center of the contours.

centring: good, average, mediocre, indifferent

The symbolic parameter profile-concavity

This parameter precises if the concavity of the curve of the projected profile is great or not.

profile-concavity: great, average, null

The symbolic parameter valiaity

This important parameter specifies the quality of the observation and thus the quality of the measured parameters; it is function of the size of the image of the galaxy. the values are:

validity: good, bad

We have seen in the previous section that for each processed galaxy five contours are built; these contours are described with both symbolic and quantified parameters. The quantified parameters correspond to the parameters which have been extracted from each contour: center-err, ellipse-err, compacity, angle, ellipticity.

Two symbolic parameters describe also the contours in the different regions of the galaxy.

The symbolic parameter contour c_i: shape

This parameter qualifies the shape of each contour, in the same way as the symbolic parameter shape; it takes the same values as the global parameter:

contour c_i: shape: elliptical, average, spiral

The symbolic parameter contour c_i: isophotes

This parameter indicates the degree of pertubation of each contour; like the global parameter associated to the whole galaxy it can take the values:

contour c_i: isophotes: smooth, normal, distorded

The symbolic parameter contour c_i: flatness

This parameter indicates if the ellipticity of each contour is not too great

contour c_i: flatness: valid

5.2.2. The prototypes

The prototypes are frames-like objects, used to define the various classes; they are organized in a hierarchy which strictly reflects the hierarchy of the classes (figure 2b). The descriptors of a prototype characterize the objects belonging to the corresponding class. Symbolic, numeric and complex structures are represented by different data structures. Some descriptors may have complex structures as shown in the following example:

Class contour
compacity: [0, 10000];
ellipticity: [-0.9, 1];
angle: [-90, 90];
ellipse-err: [0, 10000];
center-err: [0, 10000];
shape: elliptical, average, spiral;
isophotes: unknown;
flatness: unknown.

A branch of the tree of the prototypes is displayed below (the prototype Galaxy is the root node of the tree):

Class Galaxy
H: unknown;
T: [-5, 10];
shape: unknown;
bar: unknown;
profile: [0,10000];
ellipticity: [-0.9, 1];
orientation: [-90, 90];
isophotes: unknown;
linear-err: [0, 100000];
area: [0,900000];
flatness: unknown;
profile-concavity: unknown;
centring: unknown;
arms: unknown;
bulge: unknown;
validity: unknown;
c1: Class contour;
c2: Class contour;
c3: Class contour;
c4: Class contour;
c5: Class contour.

Class S upperclass Galaxy
H: S;
T: [3,7];
shape: spiral;
validity: good.

Class SB upperclass S
H: SB;
bar: present.

Class SB7 upperclass SB
H: SB7;
T: [7,7];
centring: indifferent
arms: late.

5.2.3. The facts

The facts correspond to data particular for the current case and are temporarily added to the knowledge base; for our application, the facts represent the information associated to one galaxy. The base of facts consists only of the object to be classified. This object has the same structure as the root prototype. At the beginning, all the values of its descriptors are unknown, but those corresponding to the measured parameters. At the end of the procedure, the symbolic descriptors of the object have the same values as the prototype representing the class of the object.

5.2.4. The rules

The rules are composed of three parts: the conditions, the actions and the comments. They represent the operating knowledge of the expert. The rules are used to affect a symbolic value to a descriptor of the object. In order to decrease the number of rules to scan, the knowledge base is automatically structured; a procedure of initialization attaches a few rules to each prototype of the tree. The attachment is performed if the field action of the rule operates on a descriptor of the prototype.

Example of rule:

Rule 8:

IF

ellipticity of contour3 >> ellipticity 0.1

ellipticity of contour3 > 0.4

THEN

bar be present

"As the ellipticity of the third contour is greater than the global ellipticity of the galaxy and is important, the bar is present."

5.3. The control structure

The control is only guided by the prototypes. The object, which is a priori only described by the measured parameters, is associated to the root prototype Galaxy. If no descriptor is incompatible with this prototype, then the root

prototype is considered as the current prototype. According to the context, the current prototype moves down the prototype tree.

All the successors of a current prototype which are consistent with the present values of the object descriptors are taken as hypothesis. We want to saturate the base of facts, which means to build a complete description of the object to classify; so, all the hypothesis are considered. But, as we want a natural reasoning, the prototype tree is scanned in a depth-first way (all the inferences of an hypothesis are considered before studying another one). Once, a prototype is taken as a possible hypothesis, the associated base of rules is scanned; the action fields of the activated rules increase the description of the object. Then, the prototype is validated if the distance between the object and the prototype is negligible. This distance is defined by the sum of the intermediate distances between the corresponding descriptors.

5.4. Example of working session

measured parameters:
orientation : 14.6
ellipticity: 0.50
linear-err : 0.031
profile : 0.85
area : 4300.8

contours:	center-err	ellipse-err	compacity	angle	ellipticity
c1 :	0	0.15	1.3	6.5	0.46
c2 :	0	0.05	1.9	1.5	0.31
c3 :	4	0.13	4.1	6.2	0.46
c4 :	3	0.07	4.0	5.7	0.58
c5 :	7	0.11	6.1	2.8	0.40

Activated rules: 44; 47; 48; 53; 54; 57; 61; 59; 30; 42;

selection of the prototypes: Intermediate

Activation of the prototype Intermediate
Activated rules: 4; 27;
Valididation of the prototype Intermediate

Selection of the prototypes: Intermediate–A Intermediate–B

Activation of the prototype Intermediate–B
Activated rules: 84; 89; 90;
Validation of the prototype Intermediate–B

Selection of the prototypes: LB+
Activation of the prototype LB+
Validation of the prototype LB+

Activation of the prototype Intermediate–A
Activated rules:
Validation of the prototype Intermediate–A

Selection of the prototypes: LA+
Activation of the prototype LA+
Validation of the prototype LA+

Data interpretation:

44:
As the distance between the fifth contour and an ellipse
is medium ($0.1 < d < 0.2$), this contour has an average shape.
47:
As the distance between the first contour and an ellipse
is medium ($0.1 < d < 0.2$), this contour has an average shape.
48:
As the distance between the second contour and an ellipse
is very small (< 0.1), the shape of this contour is elliptical.
53:
As the distance between the third contour and an ellipse
is medium ($0.1 < d < 0.2$), this contour has an average shape.
54:
As the distance between the fourth contour and an ellipse
is very small (< 0.1), the shape of this contour is elliptical.
57:
As the area of the galaxy is large, the measured
parameters are reliable.
61:
As the shape of the third contour is average and as the shape of
contour 2 and 4 is elliptical, the shape of the galaxy is average.

59:
As the projected profile is not linear the bulge is visible.
30:
As the second contour is circular, the orientation of its principal axis is irrelevant.
42:
As the distance between the fifth contour and an ellipse is very small (< 0.1), the shape of this contour is elliptical.
4:
As the principal axis of the third contour is perceptibly different of the galaxy principal axis, the bar is probably present.
27:
As the contours have coarsly the same principal axis and ellipticity than the galaxy, the bar is probably missing.

84:
As the shape of galaxy is average and as the shape of the fourth contour is elliptical, the arms are absent.
89:
As the ratio of the errors made by the approximation of the projected profile by a r 1/4 function and by a linear function is between 0.5 and 1, its concavity is average.
90:
As the differences between the location of the centre of the galaxy and the centers of the contours is respectively 5, 5, 5, 5 and 15 the centring is good.

-- description of galaxy 29 --
validity: good (certain)
arms: absent (certain)
shape : average (certain)
bulge: visible (certain)
flatness: unknown
profile-concavity: average (certain)
cewntring: good (certain)
bar : present (possible) absent (possible)
isophotes: unknown
class: probably classed LB+, probably classed LA+

6. RESULTS

6.1. Reliability tests

A knowledge base of 37 prototypes and 106 rules has been built. The system has been tested on a set of 21 galaxies in the Virgo cluster digitized from the same Schmidt plate. The results of the automatic classification of these galaxies, is presented in the following table:

Name	Real class	Result	Comment
NGC4459	LAR+	LA+	correct
NGC4569	SXT2	SX3	correct
NGC4474	L...P*	LA+/LB+	correct
NGC4473	E.5..	E4	correct
NGC4477	LBS.*$	L	correct
NGC4571	SAR7	SA7	correct
NGC4468	LA...*	bad validity	small area
NGC4531	S..1	LA+	slight difference
NGC4419	SBS0	SB1/SA1	almost correct
IC3392	SA.3	SB1	almost correct (dust lanes)
M91	SBT3	SB1/SB3	correct
NGC4461	LBS+*	LA/LB	almost correct
NGC4438	SAS0	LB/SB1	correct (dust lanes)
NGC4421	SBS0	LB/SB1	correct
IC800	SBT4P$	SB5	correct
NGC4523	SBS8	SB	irregular
NGC4595	SXT3$	SA1/SA3	correct
NGC4639	SXT4	SB3	correct
NGC4540	SXT6	SB7	correct
M86	.E.3..	E3	correct
M84	.E.1..	E0	correct

Although the results are quiet good, it remains a problem of false detection of a bar in presence of transversal dust lanes.

Tests have shown that the method needs a minimal size for the image of the galaxy (about 50×50). So, the rule:

if area < 2500 **then** validity bad

is activated to prevent a misclassification; in order to optimize the computing time, the parameter validity is set to good for each successor node of the root node Galaxy; so, if the previous rule is activated, no successor may be validated and the classification returns class unknown whatever are the other values of the measured parameters.

The knowledge base does not yiet contain information about the irregular, the peculiar galaxies and about the ring pattern structures. This lack of information is mainly due to the absence of such galaxies in the sampling set, so some study remains in this domain.

A second set of galaxies, with a higher resolution and various exposure time, observed by G. Paturel at the Observatoire de Haute Provence has been processed by this method; the results of the classification are displayed in the following table:

Name	Type	Result	Comment
NGC2535	4	SB7	slight difference
NGC7757	5	SB7	slight difference
IC1303	5	SB7	slight difference
NGC753	4	SA3	correct
NGC926	4	SB5	correct
NGC5371	4	SB7	slight difference
NGC7448	4	SB3	correct
NGC514	5	SB7	slight difference
NGC877	4	SB7	slight difference
NGC3992	4	SB7	slight difference
NGC2619	5	SB1/LB+	short exposure time
NGC3780	5	SB3/SB7	almost correct
NGC4254	5	SA7	slight difference
NGC5364	4	SB1/LB+	short exposure time
NGC5970	5	SB1	short exposure time

In the following table, we present the influence of the exposure time on the classification of a spiral galaxy:

Name	Exposure time	Real class	Result	Comment
NGC2336	40mn	SB4	S	lacking in precision (fig. 14a)
NGC2336	1h40mn	SB4	SB5	correct (fig. 14b)
NGC2336	3h	SB4	SB7(probably)	too late (fig. 14c)

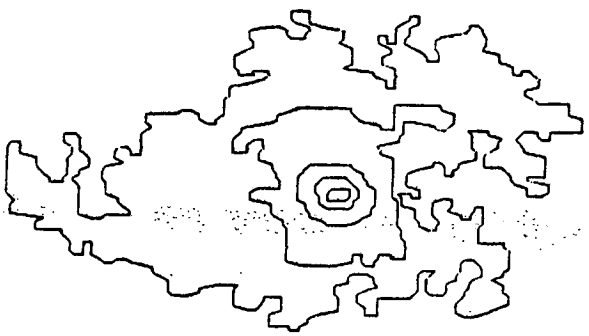

figure 14a
NGC2336 time:40mn

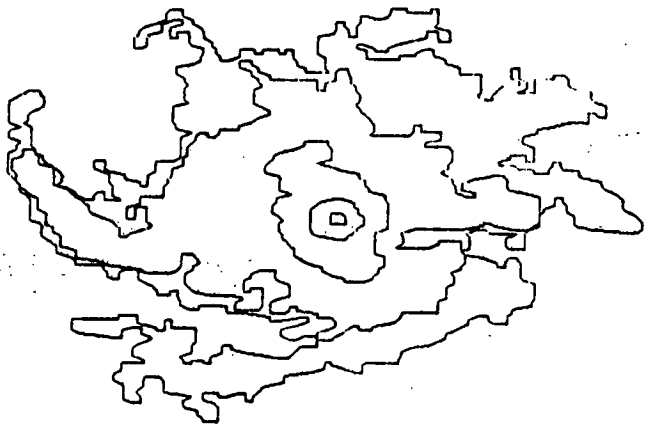

figure 14b
NGC2336 time:1h40mn

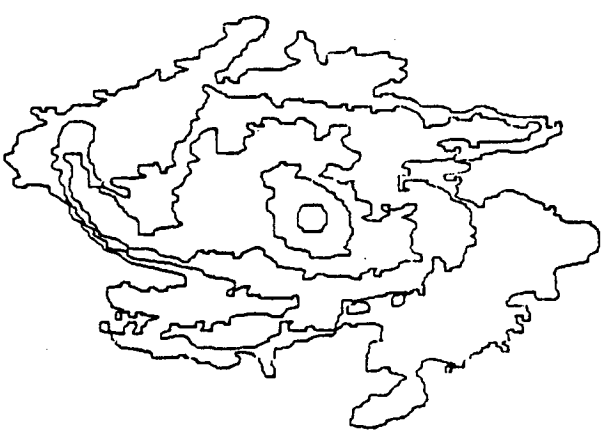

figure 14c
NGC2336 time:3h

The results obtained on a set of galaxies observed with another telescope (Schmidt telescope of Observatoire de Haute-Provence) and with different exposure times show that the knowledge base must contain information about these symbolic parameters (telescope and exposure-time). In function of the values of these parameters different branches of the tree of the prototypes must be scanned.

7. CONCLUSION

We have presented a complete automatic method to classify complex objects (galaxies), using an expert system approach.

Although the knowledge base made of 37 prototypes and 106 rules must be completed to enhance the precision, the classes obtained by this method are quiet good.

In its current state the system classify a galaxy in approximately 180 seconds: 120 seconds for the extraction of the parameters (which is of course highly dependent on the size of the image) and 60 seconds for the inference phase.

References

[**Bal82a**] D.H. Ballard and C.M. Brown, *Computer vision*, Prentice-Hall,Inc. (1982).

[**Bij81a**] A. Bijaoui, *Image et Information*, Masson (1981).

[**Bro81a**] R. A. Brooks, "Symbolic reasoning among 3-D objects and 2-D models," *AI Journal* **16** (1981).

[**Cha84a**] J. Chailloux, M. Devin, and J. M. Hullot, "LeLisp, a portable and efficient lisp system," 5^{th} ACM Conference on lisp and functional programming (August, 1984).

[**Cip84a**] P. Cipiere, *Manuel de reference du logiciel INRIMAGE*, INRIA (1984).

[**Gra84a**] C. Granger, M. Thonnat, and P. Vignard, "Etude d'un système expert pour la classification de galaxies: SYGAL," *Journees d'étude sur les systèmes experts et leurs applications* (1984).

[**Gra85a**] C. Granger, "A knowledge-based approach to object classification," *Proc of the 9th IJCAI*, Submitted (1985).

[**Han78a**] A. Hanson and E. Riseman, "VISIONS: A Computer System for Interpreting Scenes," *Computer Vision Systems*, pp.303-333, A. Hanson and E. Riseman, eds, Academic Press (1978).

[**Hay83a**] F. Hayes-Roth, D. A. Waterman, and D. B. Lenat, *Building expert systems*, Reading, Mass: Addison-Wesley (1983).

[**Hub36a**] E. P. Hubble, *The Realm of Nebulae*, Oxford Univ. Press. (1936).

[**Hul83a**] J. M. Hullot, "Ceyx, a multiformalism programming environment," *IFIP83*, R. E. A. Mason (ed), North Holland (Paris, 1983).

[**Jar81a**] J.F. Jarvis and J.A. Tyson, "Focas:Faint Object Classification and Analysis System," *Astron. J.* **86**(3), p.476 (1981).

[**Lau84a**] A. Lauberts and E. A. Valentijn, "Automatic parameter extraction for the 16000 galaxies in the ESO/Uppsala Catalogue," *The*

Messenger of ESO (1984).

[Nil82a] N.J. Nilsson, *Principles of Artificial Intelligence*, Springer-Verlag (1982).

[Pav77a] T. Pavlidis, *Structural pattern recognition*, Springer-Verlag (1977).

[San61a] A. Sandage, *The Hubble Atlas of Galaxies*, Carnegie Institution of Washington (1961).

[Sho76a] E. H. Shortliffe, *Computer-based medical consultation: MYCIN*, New York: American-Elsevier (1976).

[Slo79a] K. R. Sloan and R. Bajcsy, *World Model Driven Recognition of Outdoors Scenes*, September 1979.

[Tho84a] M. Thonnat and M. Berthod, "Automatic classification of galaxies into morphological types," *Seventh Int. Conf. on Pattern Recognition*, p.844 (August 1984).

[Tho85a] M. Thonnat, C. Granger, and M. Berthod, "Design of an expert system for object classification through an application to the classification of galaxies," *Proc of CVPR-85*, Submitted (1985).

[Vau59a] G. Vaucouleurs,de, "Classification and Morphology of External Galaxies," *Handbuch der Physik* 53, p.275 (1959).

[Vau76a] G. Vaucouleurs,de, *Le monde des galaxies*, Obs. de Besançon et lab. d'astronomie de la Faculté des Sciences (1976).

[Wat82a] M. Watanabe, K. Kodaira, and S. Okamura, "Digital surface photometry of galaxies toward a quantitative classification," *Astrophys. J. Suppl. Series* (1982).

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique